

Surgical Guidance in Pars Plana Vitrectomy Using Deep Learning and Computer Vision

Yannek Leiderman, MD, PhD

Director, Vitreoretinal Microsurgery Lab

Associate Professor of Ophthalmology

University of Illinois at Chicago

Retina Service, Illinois Eye and Ear Infirmary



Disclosures

- Alcon (C,S)
- Allergan (C)
- Genentech (C)
- Nvidia (S)
- Regeneron (C)

Summary

- We developed an automated pipeline capable of tracking the location of the surgical instruments relative to the retina during vitreoretinal surgery in real-time using deep learning and computer vision

Artificial Intelligence

- The use of machine-based intelligent agents that *mimic* cognitive functions that we associate with the human mind, such as learning and problem solving



Computer Vision

- A field of artificial intelligence that trains computers to interpret and understand the visual world
- Via digital images and deep learning models, machines can accurately identify and classify objects — then react to what they “see.”



Computer Vision

How computer vision works

Computer vision works in three basic steps:



Acquiring an image

Images, even large sets, can be acquired in real-time through video, photos or 3D technology for analysis.



Processing the image

Deep learning models automate much of this process, but the models are often trained by first being fed thousands of labeled or pre-identified images.



Understanding the image

The final step is the interpretative step, where an object is identified or classified.



Computer Vision Tools

- Image segmentation - partitions an image into multiple regions or pieces to be examined separately.
- Object detection - identifies a specific object in an image. Advanced object detection recognizes many objects in a single image
- Facial recognition - an advanced type of object detection that not only recognizes a human face in an image, but identifies a specific individual.
- Edge detection - a technique used to identify the outside edge of an object or landscape to better identify what is in the image.
- Image classification - groups images into different categories.



Deep Learning

- **Deep learning** is a subset of machine learning in AI that has networks capable of *unsupervised* learning from data that is unstructured or unlabeled.

Constraints in Retinal Surgery

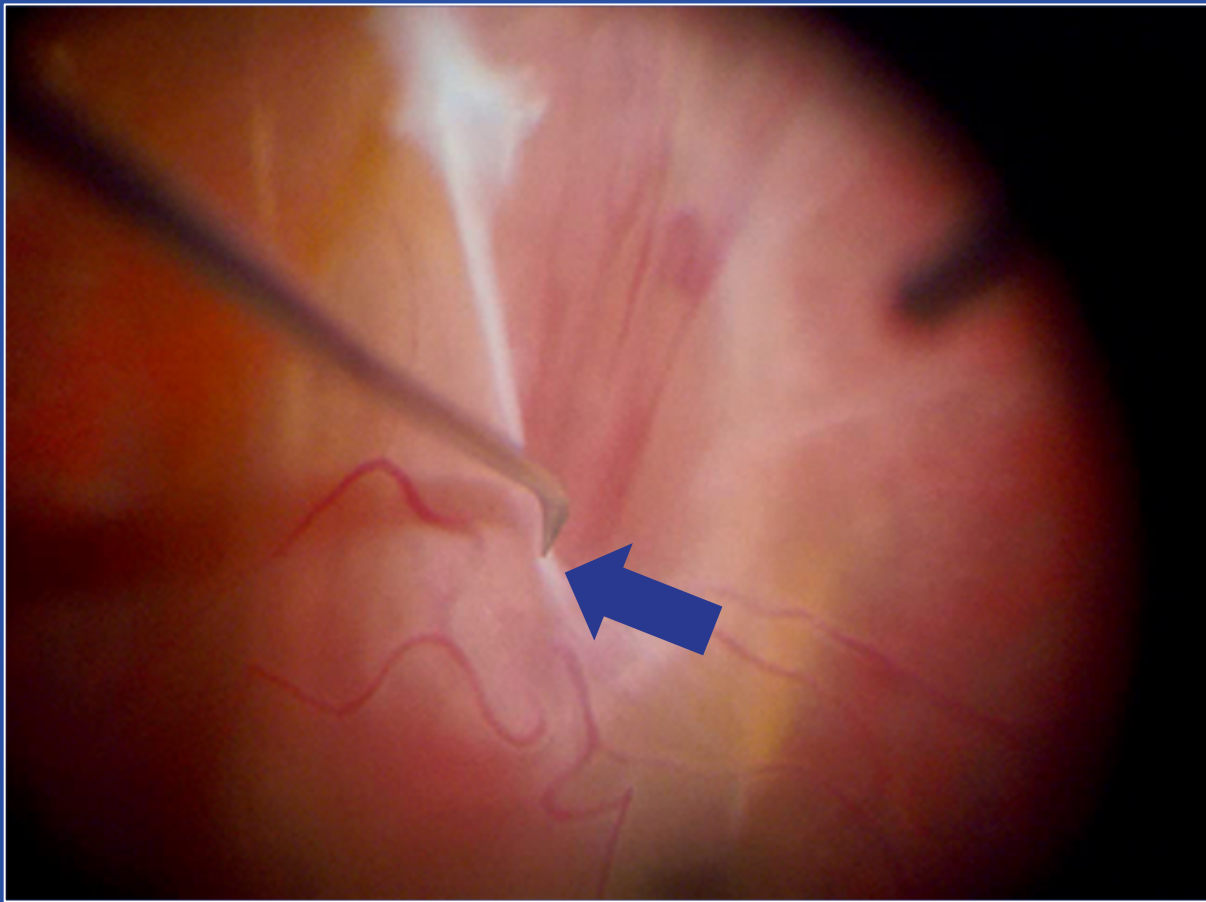
- Indirect visualization of surgical anatomy
- limited working space
- spherical, complex 3D surgical space
- instrument movement is constrained
- delicate tissues
- significant consequences to visual function from tissue damage (sight loss)



Goal

Track the position and movement of vitreoretinal surgical instruments in 3D space relative to the position of the retina to facilitate real-time, intraoperative surgical guidance during pars plana vitrectomy.





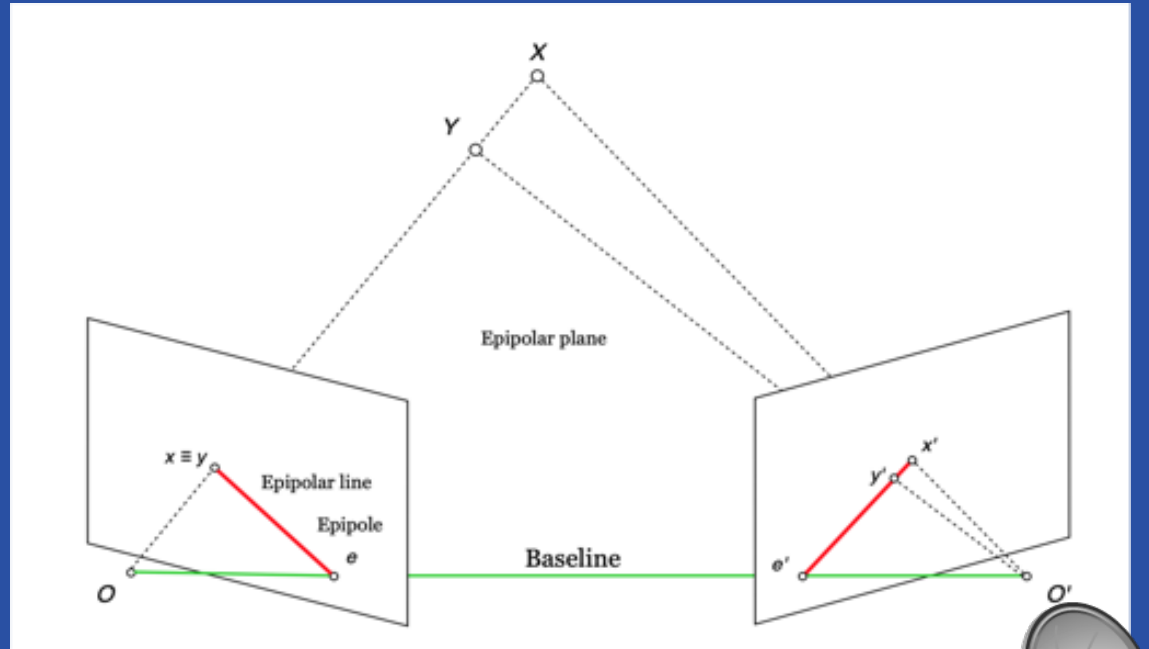
Our Solution

Implement a framework employing Computer Vision and Deep Neural Networks



COMPUTER VISION - Stereo Vision

Epipolar geometry



COMPUTER VISION - Stereo Vision (cont'd)

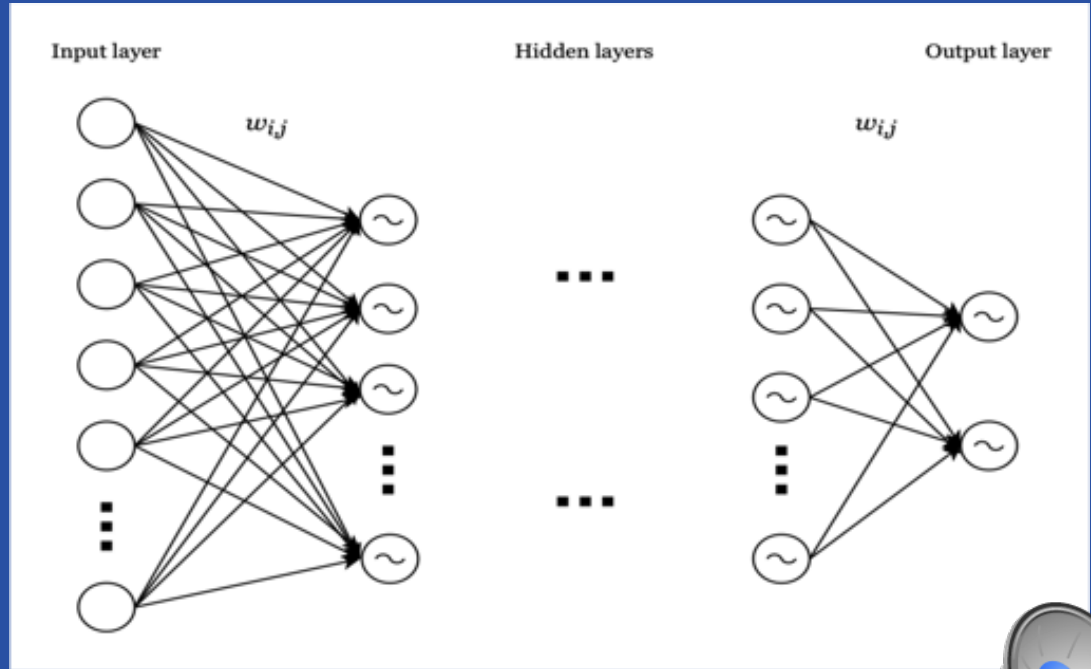
Epipolar geometry

Disparity maps



DEEP LEARNING - Convolutional Neural Networks

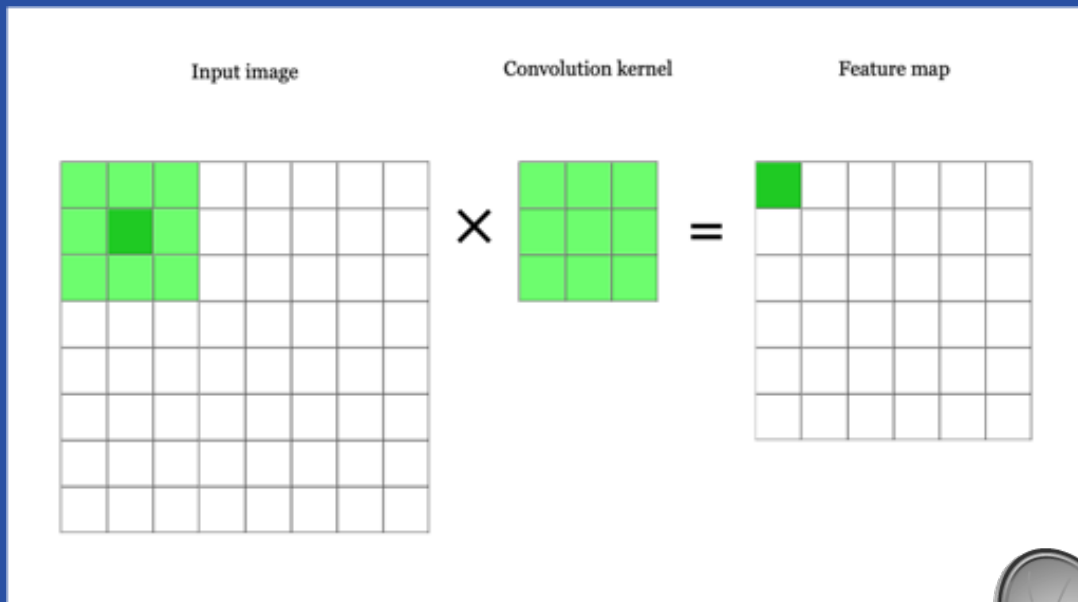
What are Artificial
Neural Networks (ANN)?



DEEP LEARNING - Convolutional Neural Networks (cont'd)

What are Artificial Neural Networks (ANN)?

What is Convolution?

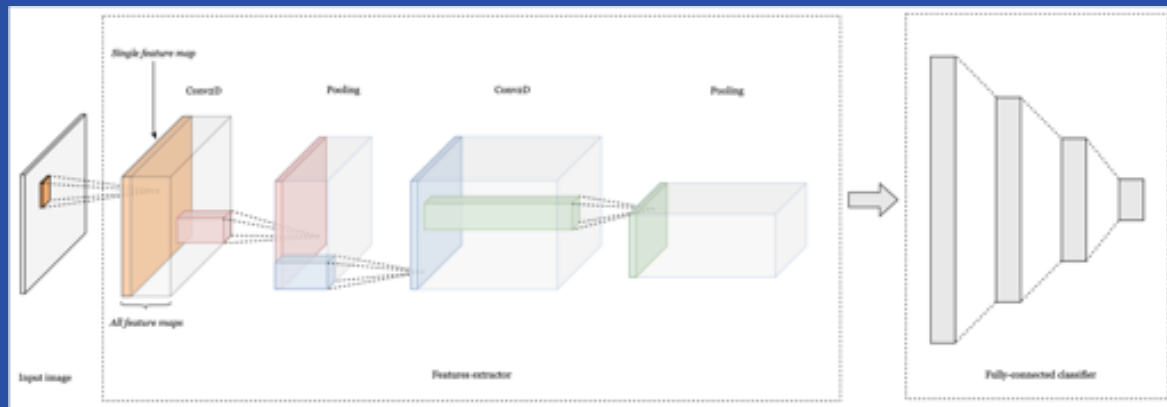


DEEP LEARNING - Convolutional Neural Networks (cont'd)

What are Artificial Neural Networks (ANN)?

What is Convolution?

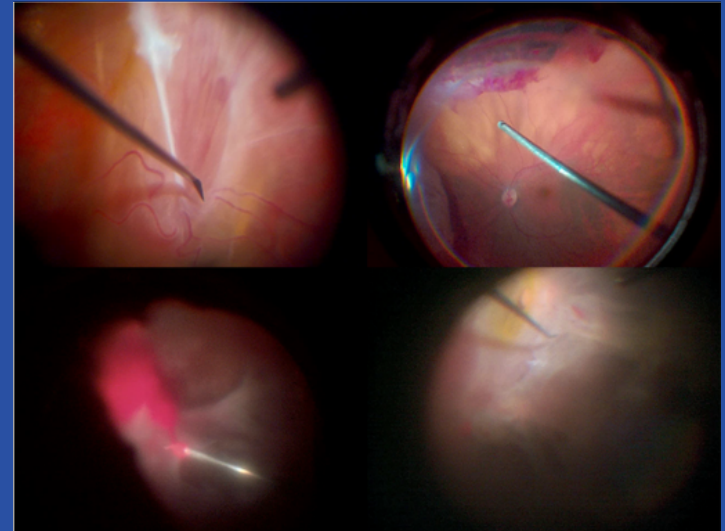
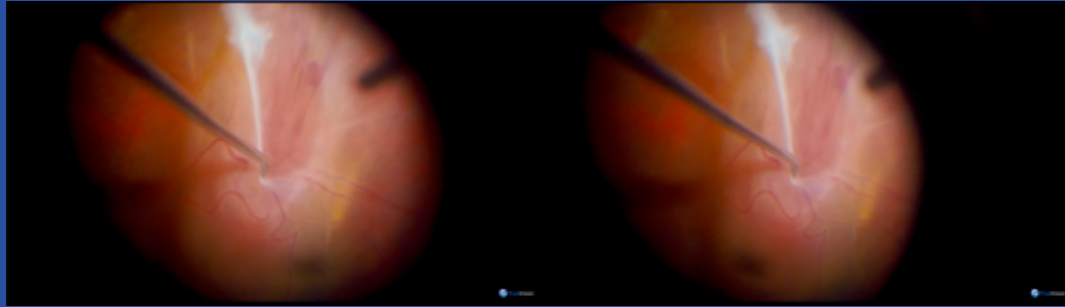
What are Convolutional Neural Networks (CNN)?



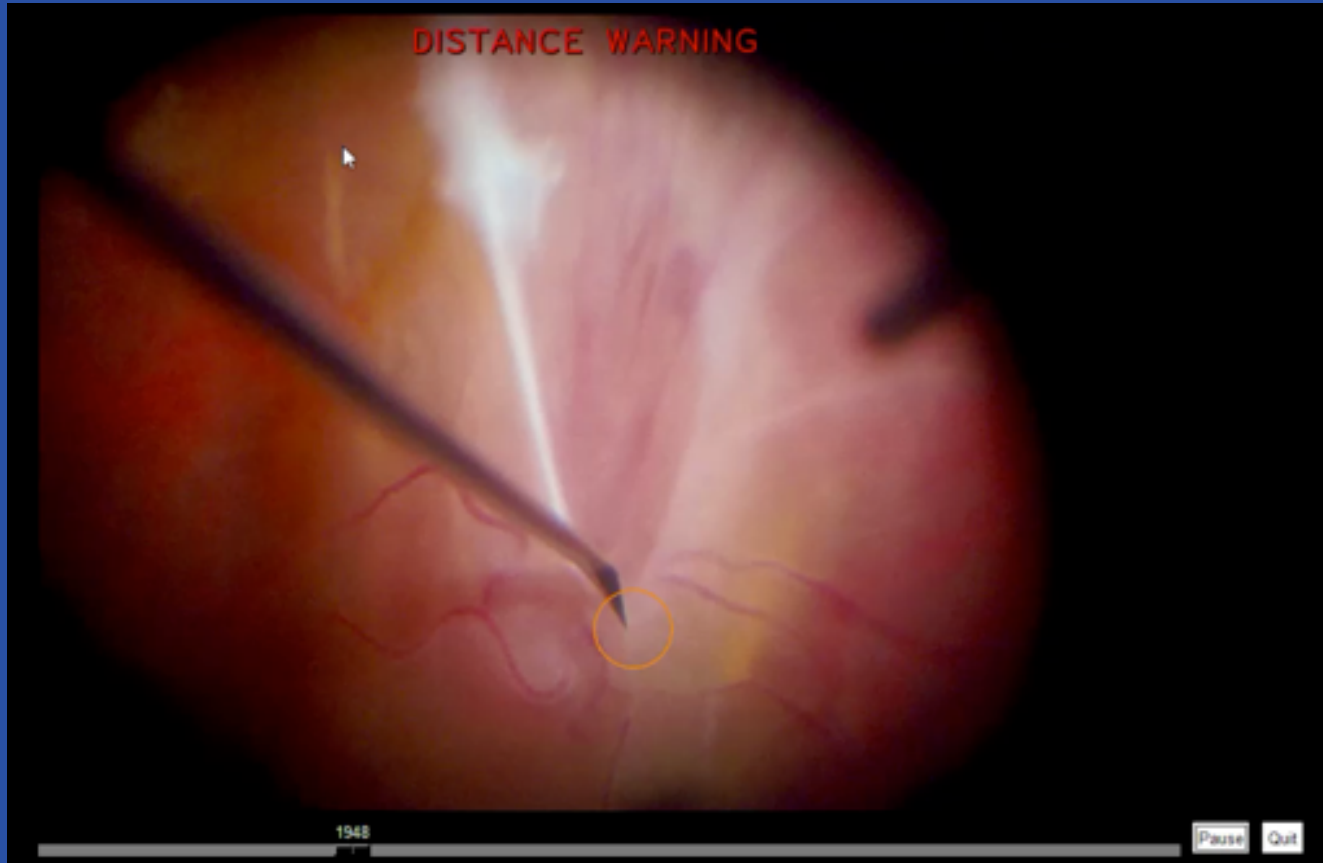
Implementation: Surgical Guidance



Pipeline - Raw Data & Training Dataset



Results - User Interface

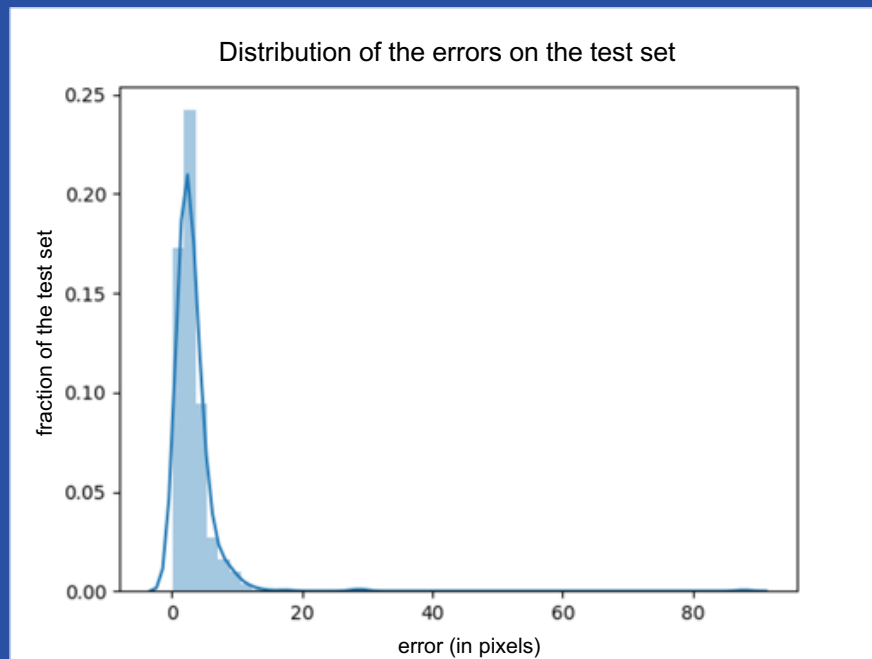


Results (cont'd)

Convolutional Neural Network performance:

- Mean error in validation phase = 3.24 pixels \pm 4.17
- Max and min errors = 87.81 and 0.11 pixels, respectively
- 90% of samples in the test sets exhibited an error of ≤ 5.53 pixels

Count	Mean	Std Dev	Min	25%	50%	75%	90%	Max
657	3.24	4.17	0.11	1.72	2.57	3.78	5.53	87.81

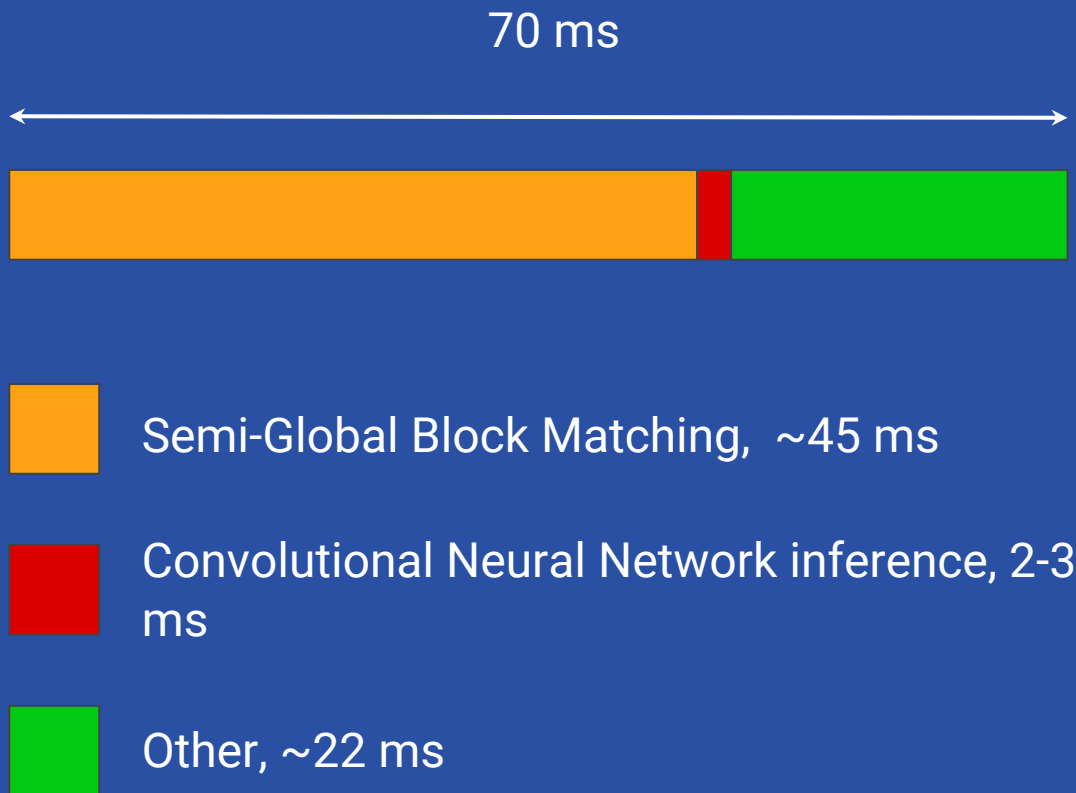


Results (cont'd)

Convolutional Neural
Network performance

Stereo-matching
performance

Pipeline performance:
Single iteration of the pipeline
required 70 – 100
milliseconds



Conclusions

Proof of concept for:

Effective instrument tip
localization with CNNs



Computed distance
between instrument and
retina from 2D images



Achieved real-time
performance



Implemented framework for
surgery guidance



Summary

Computer vision tools and deep learning have been applied to static retinal imaging

- Aid in clinical image interpretation
- More recently in clinical decision making



Summary

- Application of AI in real time to dynamic imaging used in intraoperative digital surgical visualization can be used for tool- and tissue tracking in surgical guidance



DISTANCE WARNING



Tanya Berger-Wolf PhD

Cristian Luciano PhD



Wei Tang PhD

Mattia Di Fatta

Marco Santambrogio PhD

